

<https://helda.helsinki.fi>

Recycling a genre for news automation : The production of Valtteri the Election Bot

Haapanen, Lauri

2020-10-07

Haapanen , L & Leppänen , L 2020 , ' Recycling a genre for news automation : The
production of Valtteri the Election Bot ' , AILA Review , vol. 33 , no. 1 , pp. 67-85 . <https://doi.org/10.1075/aila.00030>.

<http://hdl.handle.net/10138/321232>

<https://doi.org/10.1075/aila.00030.haa>

cc_by_nc

publishedVersion

Downloaded from Helda, University of Helsinki institutional repository.

This is an electronic reprint of the original article.

This reprint may differ from the original in pagination and typographic detail.

Please cite the original version.

Recycling a genre for news automation

The production of Valtteri the Election Bot

Lauri Haapanen & Leo Leppänen

University of Jyväskylä | University of Helsinki

The amount of available digital data is increasing at a tremendous rate. These data, however, are of limited use unless converted into a user-friendly form. We took on this task and built a natural language generation (NLG) driven system that generates journalistic news stories about elections without human intervention. In this paper, after presenting an overview of state-of-the-art technologies in NLG, we explain systematically how we identified and then recontextualized the determinant aspects of the genre of an online news story in the algorithm of our NLG software. In the discussion, we introduce the key results of a user test we carried out and some improvements that these results suggest. Then, after relating the news items that our NLG system generates to general aspects of genres and their evolution, we conclude by questioning the idea that journalistic NLG systems should mimic journalism written by humans. Instead, we suggest that developmental work in the field of news automation should aim to create a new genre based on the inherent strengths of NLG. Finally, we present a few suggestions as to what this genre could include.

Keywords: news automation, natural language generation, NLG, genre, journalism

1. Introduction: Converting large data sets into user-friendly form

There is more data available than ever before; the amount is estimated to double every 40 months (Latar, 2015). There is also a tendency in modern societies toward opening up publicly financed data sets. However, this openness remains largely theoretical. On the one hand, journalistic media with their dwindling resources can hardly sift through such data and disseminate relevant parts of it very extensively. For individuals to leverage these data sets requires advanced technological skills and significant effort. Therefore, in order to take advantage of this free information and make it contribute to public discussion, we – the authors together with

other researchers in the project called Immersive Automation¹ – decided to see how natural language generation (NLG) could respond to this dilemma.

What this meant in practice was that we built a computer program that automatically converts relevant parts of a large data set into a user-friendly format as a journalistic news item. Example (1) demonstrates the type of text produced by our system. The example deals with the Finnish municipal elections that took place in April 2017. It focuses particularly on the results for Padasjoki, which is a small municipality in Southern Finland and, therefore, hardly worth any exclusive coverage by human journalists. Our NLG software, however, “wrote” the news item without human intervention in a fraction of a second, and it could do the same for any region, municipality, polling station, party, or candidate – reporting whatever the reader wants to know.

(1) News item “Padasjoki”

Biggest gains for the Green Party in Padasjoki

The Green Party increased their number of seats the most in Padasjoki and secured 4 more seats. The party is the largest party in the council and got 17.2% more votes than in the last municipal election. The party increased its voter support by the greatest margin and has 7 seats in the new council.

[Two paragraphs removed.]

Ari Rantanen (Green) received most votes and secured 5.4% of the vote. He got 94 votes and was elected to the council.

The production process resulting in the NLG system that generated this piece about Padasjoki consisted of separate instances of *recontextualization* (Linell, 1998). Starting from a tabula rasa, our research and development team first closely analyzed the genre of journalistic news in order to identify what aspects or features are the determinants of a text that would be representative of this genre, and what genre-specific social activities would lead to this outcome. Such aspects could be assessments, values and ideologies as well as ways of seeing and saying things and acting toward them (1998, pp. 154–155). Then we encapsulated and extracted, one by one, these aspects and repositioned them through a process of “dynamic transfer-and-transformation” (1998, p. 154) in our emerging algorithm. In the next sections, after presenting an overview of state-of-the-art technologies in NLG, we will introduce

1. The project was conducted in 2017–2018 by researchers from the University of Helsinki and the VTT Technical Research Centre of Finland. It created a roadmap and a demonstration of a news ecosystem based on news automation and intense audience engagement. The experimental NLG software system discussed in this paper was created within this project. For more information, see www.immersiveautomation.com.

the genre in question and explain systematically the series of recontextualizations that we went through in the production of the algorithm.

In the discussion, we introduce the key results of a user test and some improvements that could be made on the basis of what we found out there. After relating the news items that our NLG system generates to general aspects of genres and their evolution, we challenge the idea that an NLG-driven system mimics journalism written by humans and argue for the deliberate creation of a new genre that is based on the inherent strengths of NLG. The paper ends with a preliminary discussion of what the features of this new genre might be.

2. State-of-the-art in NLG: A spectrum from ruled-based to training-driven methods

In this paper, we define natural language generation as a process that takes non-linguistic input, such as a database of numbers or an image, and produces textual output in a natural language by following instructions coded in an algorithm (Gatt & Krahmer, 2017). Traditionally, this process has been seen as consisting of multiple subproblems, most famously as presented by Reiter and Dale (2000):

1. **Content determination:** deciding what information, i.e., which set of messages from the system's inputs or underlying data sources should be communicated in the text.
2. **Document structuring:** imposing order and structure over the set of messages to be conveyed.
3. **Sentence aggregation:** grouping messages together into sentences.
4. **Lexicalization:** deciding which specific words and phrases should be chosen to express the domain concepts and relations which appear in the messages (e.g., is the event represented in the DEPARTURE message expressed with the word *leave* or *depart*?)
5. **Generation of referring expressions:** selecting words or phrases to identify domain entities (e.g., using referring expressions *a party* and *it* to refer to the domain entity POLITICAL-PARTY).
6. **Linguistic and surface realization:** applying the rules of grammar to produce a syntactically, morphologically and orthographically correct text.

Gatt and Krahmer (2017) note that while this series of tasks is useful from a conceptual point of view, many real-world NLG systems conduct these tasks in a partially overlapping way while others further divide certain tasks into smaller subtasks. In fact, some systems, most prominently those based on neural networks, avoid this division into tasks entirely and instead view NLG as a unified, global process. For the purposes of our analysis, we will consider the above subtasks as grouped into

three larger tasks, namely, *document planning* (#1–2 above), *microplanning* (#3–5 above) and *realization* (#6 above) (cf. Diakopoulos, 2019, p. 98).

For the purposes of the analysis described below, it is also relevant that we are analyzing the process of building an NLG system based on rules written by humans. That is, our system consists of source code that explicitly defines what decisions the system is to make, as well as how those decisions are to be made. Recent studies on NLG have, however, explored entirely different approaches as well, and the systems described in the literature form a spectrum from systems based on rules written by humans to those based on trained, data-driven methods (Gatt & Krahmer, 2017).

The trained approaches are perhaps best exemplified by systems that employ deep learning methods. These machine-learning systems typically ingest training data in terms of pairs of input and expected output. Based on this training data, the system then attempts to learn to mimic the process that created the training data. While such an approach is in many ways appealing, in the current state-of-the-art it is difficult to have full control over the actions of the model, which makes these systems less appealing for the many genres where correctness is paramount over textual fluency and variety of expression.

One such problem is “overfitting” (e.g., Murphy, 2012), where a model performs well on the training data but fails to generalize to cases outside the training data. A machine-learning system making hiring decisions might notice that – according to the training data – no previous candidate hired by the company was born on a Tuesday. To a human, this is obviously irrelevant, but a computer might create a model that rejects all future candidates who were born on a Tuesday. In NLG, this problem surfaces as the system “hallucinating” content. For example, a system for producing restaurant reviews might learn that restaurants of a certain type are almost always inexpensive, resulting in reviews where similar restaurants are characterized as inexpensive irrespective of their real price points.

Journalism, the focus of this paper, requires high accuracy and accountability. This paper therefore adopts a rule-based NLG method over a method based on machine learning.

3. Case study: The production of Valtteri the Election Bot

The aim of our NLG software was to identify relevant pieces of information, i.e., *facts*, stored in a data set and then disseminate them to a diverse audience in a user-friendly format. To systematically achieve this aim, an obvious vehicle was the genre of *journalistic news*. While the specific genre that we were aiming at is often referred to as *breaking news*, *a news brief*, or *hard news*, we decided to use *an (online) news story* – that is interchangeably, with or without specification of the

exact medium. Whatever the exact nametag, it can be considered a genre – a conventionalized, language-mediated means of performing intended social actions – because all the articles share certain characteristics: the various (online) news stories all share more or less the same aim, participant roles, medium of communication, and linguistic structure, i.e., a so-called inverted pyramid style (Miller, 1984). The inverted pyramid conveys the most important information in the first paragraphs and then goes on to other material in descending order of importance; the organization “resulted from the professional effort to strengthen the communicative quality of news” (Pöttker, 2003, p. 501).

The genre we had selected guided the production of our NLG system. In what follows, we will explain the research and the decisions we made. The process consisted of determining not only the details of the three technical NLG tasks, but also of a preceding *domain selection* phase. However, work on deciding the details of the phases did not occur strictly linearly: rather, the work on the different parts was iterative and overlapping. Furthermore, for the sake of this article, we have cut corners and left out some details, as the focus of this paper is on the process of recycling a genre; this paper is not intended as a complete description of how to put together an NLG system. Similarly, the details of programming fall outside the scope of this article (for technical details, see Leppänen et al., 2017a, 2017b).

3.1 Domain selection: Topicality and relevance

In the first phase of the production, we needed a **domain** that our online news stories would discuss. As news prototypically deals with something topical as well as relevant from the audience’s point of view, we looked for a domain that meets these requirements. At the same time, we had to keep in mind the formal requirements that successful NLG production sets for a data set, especially the need for data to be systematically structured and mostly numeric or (near-)categorical (e.g., names of people).

Given these criteria, we selected municipal elections in Finland as our domain, and named the prospective software *Valtteri Vaalibotti* / *Valtteri the Election Bot*, playing on the Finnish term for elections, *vaalit*.² These elections are organized every four years, and at the time when our project was beginning, they were a topical issue. As a large and influential political event, they are also of great relevance to Finns. The results of the elections were released by the Finnish Ministry of Justice immediately after the completion of the count, and they were provided online in a machine-readable numeric format. Last but not least, the core principle of journal-

2. Later, the system was repurposed to work with data on crime statistics and renamed *Crime Valtteri*.

ism demands that every piece of information should be true. The election data we used were produced by the Ministry of Justice, an institution that could be – at least in the Finnish context – relied on.

In detail, the data corpus provided by the Ministry included the results for every party on the national level, and on the levels of each of the 13 electoral districts, 311 municipalities, and 2,012 polling stations. For each of the 33,316 candidates, the data included details of their success in their municipality and in each of the municipality's polling stations. The files also included some information on the candidates' background, such as sex and party affiliation, as well as a comparison of their success in previous elections. After being preprocessed for use with the NLG software, our data set consisted of more than 10 million discrete values that could be expressed in the text. As the inverted triangle answers such questions as *who* did *what*, *why*, *when*, and *where*, we next needed to identify these pieces of numeric information from the huge data corpus in order to recontextualize them in a news story.

3.2 Document planning: Readers' preferences, news values, and rhetorical structures

The next step after selecting the domain was to determine what facts out of the data corpus would be selected for any particular news story addressed to any particular audience. Each news story – or a collective platform of news stories such as a newspaper – is always directed to a certain readership.

To begin with, a nationwide election is naturally an event of interest to a wide readership. Examined more closely, however, the interests of each individual citizen may vary. As we wanted *Valtteri* to serve as wide a readership as possible, any particular story had to take into account the **reader's preferences** and be tailor-made accordingly. In practice, we designed an online user-interface that allows the reader to select as the vantage point of a story (1) the entire country, or a specific region, municipality, or polling station, and optionally (2) a specific party or candidate. The reader could also decide if the article was to be written in (3) Finnish or Swedish, the two official languages of Finland, or in English. In the Padasjoki story (Example (1)), the selections were made with regard to (1) the geographical area (the municipality: Padasjoki) and (3) the language (English), but not with regard to (2) any party or candidate.

To select the most newsworthy issue, given the assumed interests of the readership, further criteria were needed to whittle down a large number of potential facts to just a few. In journalism written by humans, these procedures are often adopted through socialization and performed tacitly and unconsciously. However, by analyzing news products and newswriting processes, research has built

up lists of the characteristics that circumstances or events should have in order to be elevated to newsworthy status and then be constructed and mediated as news. In brief, the more **news value criteria** an event satisfies, the higher the probability that it will become news. Despite the fact that the criteria and their mutual dynamics are a subject of constant practical and scholarly debate (for an overview, see O'Neill & Harcup, 2009; Zampa, 2017, Chapter 4), we view these dozens of criteria as effectively creating four distinct larger categories, *topicality*, *unexpectedness*, *societal importance*, and *personal interest*. A significant factor in newsworthiness seems to be the *outlierness* of the events when contrasted with the set of normal, day-to-day events (for relevant literature on automatic outlier and novelty detection, see, e.g., Chandola, Banerjee, & Kumar, 2009; Gupta et al., 2014). On this basis, we came up with a method for automatically determining the overall newsworthiness of any given fact presented in the election data.

At this stage in the production, readers' preferences and news values were recontextualized as the criteria to screen out and rank the huge number of facts that the original data corpus contained. However, bluntly presenting the dozen most important facts one after another, either in random order or in order of newsworthiness, is not the most rhetorically convenient way of persuading readers and enhancing the delivery and reception of information. Therefore, in determining the rhetorical structure, we consulted the existing literature on writing research and also conducted some **rhetorical analysis** of our own to figure out the genre-specific way to organize and combine the relevant facts.

As a result, *Valtteri* was guided to build each story using three to five multisentence paragraphs, each of which should deal with only one theme. Reflecting the inverted pyramid style, the theme of the first paragraph – and its first sentence, too – was determined by the fact that it was ranked as the most newsworthy. In the Padasjoki story, the highest-ranked fact, set into an appropriate phrase/template, reads “The Green Party increased their number of seats the most in Padasjoki”. The rest of the paragraph then consists of facts around the same theme (the success of the Green Party) in descending order of newsworthiness.

After that, the second paragraph determines its theme by selecting the next most highly ranked fact that is sufficiently different from the theme of the first paragraph. In the Padasjoki story, this is the success of the National Coalition Party. In order to follow genre-specific conventions, the narration of the news stories should be progressive and the paragraphs must be relatively short, for example, in comparison with the length of paragraphs in fiction. We, therefore, determined that one paragraph should consist of no more than seven theme-bound facts. At the same time, we did not want the story to descend into minutiae, and so we instructed *Valtteri* to end the paragraph before reaching that number if there were no more sufficiently newsworthy facts that could be used in it.

In the next phase, these facts, which had been represented so far only as abstract information rather than as language, needed to be given a genre-specific linguistic description. In other words, they had to be recontextualized into the human language system.

3.3 Microplanning: Lexicalization and aggregation

Following the document-planning process, the system needs to determine how to convey the selected newsworthy facts linguistically. This process is often seen as starting off with **lexicalization**, where the words and syntactic structures of the text are determined (Reiter & Dale, 2000). In our case, experts in Finnish-, Swedish-, and English-speaking journalism wrote a template of a basic phrase – including both the general syntactic structure and the operative words, especially verbs – for each fact in natural language. In other words, they recontextualized their knowledge about the genre-specific language use in the emerging algorithm; on the one hand, about the values and ideologies (e.g., accuracy, fairness, accountability), and on the other hand, about the formal requirements (e.g., clarity, brevity, unambiguity).

From a technical point of view, these basic template phrases are the driver of the linguistic realization from which the entire texts are then generated, for example, by combining them. They, therefore, have to be flexibly interconnectable and include rules that tell the system when each of the basic phrases/templates is applicable.

Below, there are two examples of phrases/templates and their instructions. They exemplify two extremes of the types: while Example (2) is still very abstract, only containing the canned words “has” and “in”, Example (3) is close to a surface text, only lacking the name of the person.

- (2) {name} has {value} {unit} [in {place}]

Applicable to: any fact discussing a total number or percentage of seats or votes obtained by either a person or a party at a single location at a single time.

- (3) {name} is a newcomer [in the council]

Applicable to: any person who is not a member of the municipal council at the time of the election

The {curly brackets} designate a part of the phrase/template that is to be replaced by content at a later stage. The [square brackets] designate a part that can be omitted if it would constitute undue repetition.

At this point, all the facts have a genre-specific and idiomatic natural language realization in three languages. To enhance readability, *Valtteri* is allowed to combine (known as *aggregation* in the literature) two phrases/templates to form a single sentence. This is a complex, multifaceted process; even the relatively “simple”

syntactic aggregation needs to consider multiple ways of combining two phrases (Gatt & Krahmer, 2017).

- (4) The party took 524 votes and secured 30.2% of the vote.
- (5) The National Coalition Party is the largest party in the council but it got 11.3% fewer votes than in the last municipal election.
- (6) The party got 23.8% of the votes, which is 4.1% fewer votes than in 2012

For example, to decide whether to combine two clauses with the word *and* or *but/although*, *Valtteri* needs to evaluate the effect of the sentences from the point of view of the subject of the sentence. That is, *but/although* is only applicable if the clauses have a non-negative and a negative effect (see Example (5)), whereas *and* is more applicable for situations where the effects are both either negative or non-negative (Example (4)). Using more complex structures such as *which is* (Example (6)) also requires analysis of whether the phrases discuss related aspects of the same entities.

It is at this stage of the process that the system also needs to decide how to refer to various entities (i.e., people and parties) both when first mentioned and on subsequent mentions. Regarding candidates, *Valtteri* follows the genre-related conventions of neutral political reporting by using the full name of a candidate with an abbreviation of the party's name in parentheses. However, if the candidate has already been referred to, *Valtteri* leaves out the last name and the affiliation, and if the candidate is the person whom the text last referred to, *Valtteri* uses a pronoun (see Example (7)). A similar approach can be taken with parties, where the full official name is used initially, but later references use a more common, shorter version of the name, and where there are consecutive references to the same party, the phrase *the party* or the pronoun *it* are used (see Example (8)). It should be noted that in the case of the parties, the common, shorter, names themselves are a genre convention that has arisen in the local news industry.

- (7) Ari Rantanen (Green) / Rantanen / he
- (8) the Green Party / the Greens / the party / it.

In this microplanning phase, our NLG system determined how to convey the abstract information content of the story, which had been decided during the document planning, in human language. Finally, the emerging news story had to be formatted to meet genre-specific criteria in terms of orthography and typography.

3.4 Realization: Orthography and typography

In the previous section, we illustrated the process of recontextualization with some examples. They looked like a “proper” text with their capitalization and punctuation because they were extracts from the ready-made news story (Example (1)). However, in effect, the result of the text realization stage is a tree structure of text snippets, each snippet corresponding to a sentence in a ready-made news story. The higher-level structures of the tree, then, correspond to paragraphs.

The final step of the generation process is the flattening of the tree structure into a linear text in order to produce an orthographically correct text. Here, too, several genre conventions about news stories’ layout were taken on board.

Firstly, it must be observed that for applications other than journalism, the system would not necessarily have to map the constructs of the tree-form input into sentences in paragraphs. In another genre, it might be more appropriate to realize text snippets as consecutive bullet points and the higher-level structures as, for example, slides in a PowerPoint presentation rather than as paragraphs of a text. Even the decision to output what could be considered a standard text with paragraphs is thus influenced by the chosen genre.

Similarly, some other orthographical details that might be taken for granted are actually genre-related. For example, if the text was intended as a chat message, the formally correct capitalization and punctuation could – or even should – be omitted in order for it to have an authentic layout. Some decisions-by-default, so to speak, are easier to identify by carefully considering possible alternatives: in other genres, such as Tweets, the texts could include additional elements, such as #hashtags and emoji.

Finally, in the case of *Valtteri* the finished texts were going to be presented to readers on screen. With this in mind, several decisions regarding the typography were made – again, influenced by the news genre. First of all, the body text of the emerging news story was rendered in a “serious” font, Roboto (Example (9)), rather than anything out of the ordinary that could perhaps be seen as sensational or frivolous. Secondly, the heading of the text was distinguished from the body text by means of a different and larger font size (Example (10)).

(9) This Roboto font was selected to reflect the genre-specific factuality.

(10) The font Oswald was chosen to be a little more eye-catching and weightier than that of the body text, but still very much a “matter-of-fact” font.

With regard to paragraph separation, we chose to go with the web standard of separating paragraphs with whitespace rather than the newspaper standard of first-line indentations. Finally, the text width was set to a relatively small character count, as in many news websites. In all these decisions, both the genre (news as

a matter-of-fact and serious genre) as well as the publishing medium (web pages rather than physical print) were considerations. With a different genre or different medium, most of these decisions would probably have been different.³

In this and previous sections, we have introduced the research and production process by which we built our NLG system that generates online news stories without human intervention. In the next section, we will present the key findings of a user test and discuss some considerations that arose in consequence.

4. Discussion: Toward a distinctive genre of online news story

Once *Valtteri* the Election Bot was up and running, we conducted a user test to figure out how successful we had been and what could and should be improved. The user test consisted of two parts. First, evaluators ($n = 152$) were asked to rate ten preselected stories, six of which were made by *Valtteri* and four by human journalists. Then, the evaluators were asked to select and rate four more articles of their own choice using *Valtteri*'s user-interface. The aspects of evaluation were *credibility*, *liking*, *quality*, and *representativeness*. (For details of the test setup and results, see Melin et al., 2018.)

The news stories generated by *Valtteri* received statistically *significantly* lower ratings on all factors compared with those stories written by journalists. *Valtteri*'s best rating was for *credibility*, and this was the only somewhat comparable value to the scores for human-written stories. It is worth drawing attention to the fact that one of *Valtteri*'s articles got an even better rating for credibility than the corresponding journalist-written article (similarly in sports news, see Wölker & Powell, 2018). The results for *representativeness* were so-so, and those for *quality* and *liking* were the weakest. (Similarly, see Clerwall, 2014; Graefe et al., 2016; Wölker & Powell, 2018). Interestingly, those *Valtteri* stories that the evaluators could freely select got slightly better ratings than the preselected ones, statistically significantly for *liking* and *quality*.

The users were also asked to give written feedback on the stories. The most frequent complaints were about language errors, obtrusive repetition, and "dry" language, and the most common words in the negative feedback were words like *boring*, *confusing*, *monotone* and *incoherent*. On the positive side, the computer-written stories were generally praised for being based on facts and for being clear and to-the-point. *Valtteri* succeeded best at *credibility*, as mentioned above, and

3. We also produced versions of the texts with the sorts of additional visual elements that one could find in a print medium, such as portraits of the candidates, party logos, and locator maps. These visual elements, however, are left out of this analysis.

this was reflected in the most common positive words: *facts, numbers, informative, objective, equal* and *credible*.

All in all, it can be concluded that significant improvements need to be made before *Valtteri* can be used to produce ready-to-publish news stories. In what follows, we discuss the nature of *Valtteri*'s news stories and the improvements that would be necessary in the light of genre theory.

4.1 Challenge: Trade-offs to improve Valtteri's performance

The notion of *genre* in linguistics refers to the established way in which certain participants use language in a specific medium for a particular purpose in particular social, societal, and cultural contexts (e.g., Fairclough, 1992; Martin, 1985; Miller, 1984). However, genre is "only" an abstraction, in that each of its realizations displays the characteristics of the ideal type of the genre but also features individual variation, which reflects the influence of the actual communication situation (Gruber, 2019). From this, it follows that the boundaries of a genre are not clear-cut but are defined according to the degree of typicality of the characteristics. In this light, *Valtteri the Election Bot*'s activity contains a significant non-typical feature. The layout, structure, subject matter, content and narration of *Valtteri*'s stories do reflect, and contribute to, the aims of a particular linguistic social activity, which is a specific communication event between journalists, sources, and readers when a breaking news event has taken place – with one exception: an algorithm, instead of a human journalist, occupies the participant role of the producer of the news story.

It is common to describe genres from the point of view of their stable and reproducible characteristics. Genres provide stability and continuity in otherwise varying, context-dependent instances of language use. In other words, each individual genre combines similar situations of language use with each other and with the activities that these individual situations represent. (Pietikäinen & Mäntynen, 2020). Or, to formulate the dynamic from the language user's perspective, genres have to be stable in order to remain recognizable and, thus, facilitate communication by providing a reliable pattern for fulfilling a communicative need in a particular situation of language use (Luginbühl, 2014). However, none of these situations is identical, and, therefore, dynamism and flexibility, too, must be inherent characteristics of genres (Devitt, 2004). The genre of an online news story tolerates a change in the range of participant roles, otherwise the textual outcome of *Valtteri* as well as of other NLG-driven news automation systems could not be identified as representative of this genre. However, this change impacts the textual outcome in contradictory ways.

The user test confirmed that the linguistic expression of *Valtteri's* news stories leaves much to be desired. This deficiency could be addressed – at least to some degree – by means of technical improvements, such as a wider variety of templates and idiomatic expressions as well as more complex aggregation. These, however, come with their downsides. Automatically identifying which of multiple slightly semantically different phrasings is the most appropriate is extremely difficult and can easily lead to error, and this is increasingly the case as the variety of expression increases. Similarly, it is unclear whether *Valtteri* can, for example, determine the causal relations between the facts of the story with such certainty that using the connector *because* is feasible.

Furthermore, in order to improve the relevance of *Valtteri's* stories, contextual factors need to be better taken into account when evaluating the newsworthiness of individual facts. For example, even a modest vote for a celebrity or locally prominent figure would be a newsworthy issue, at least locally. However, exhaustively identifying such people automatically would be very difficult, and such decisions would probably in many cases be highly debatable. If the attempt were made, it would most likely lead to errors and unequal treatment of the different candidates – undermining something, *credibility*, that was considered in the user test to be a positive feature of *Valtteri*.

While none of the things mentioned above are impossible, they all entail more and more development work and research, both of which translate directly into time and money in a real-world setting – and both of these are things that modern newsrooms lack. This being the case, any real project to implement an NLG system must balance the quality of the system with the resources available for its development.

4.2 Solution: From non-typicality to a new genre

As we have explored above, dynamism and flexibility characterize genres in a synchronic perspective and enable genre change in a diachronic perspective. This means that genres become transformed through gradual adaptation to changes in contexts and uses over time. (Devitt, 2004)

The pursuit of speed serves as an example of such an evolutionary process. Speed has always been of the essence in journalism (e.g., Rosenberg & Feldman, 2008; Weaver & Willnat, 2012), and the emphasis on it has further increased as a result of digitalization and the Internet (e.g., Lee, 2014). NLG-driven news automation has taken temporality to extremes. *Valtteri* produces news stories in a split second; in comparison, it took some 60 minutes for human journalists to write one control story for the user test. In this regard, one branch of the genre of news, which we have named online news stories, has evolved enormously over

recent decades. It may therefore be reasonable to argue that it has already entered the gray area where non-typicality transforms into a new genre.

In addition to speed, another unique feature of *Valtteri* is advanced personalization. It is true that targeting content to a specific, even niche audience is prototypical of media outlets in general, but our research and development team aimed at a categorical shift: *Valtteri* is not gradually narrowing down the target audience but is starting the entire text production from a single reader's interests.⁴ This change of approach affects the news value criteria: instead of finding an angle that many people will find interesting, the same topic can be approached from numerous angles, depending on who is reading the article. This change can be considered more than the evolution of an existing genre; it is a step toward a new, categorically more personalized news genre. At the same time, it shows that, instead of in spontaneous evolution, generic change can also originate in a conscious effort to take advantage of new technical affordances (Miller & Shepherd, 2009).

The decision to focus on personalization proved to be a good one. In the user test, as we have already mentioned, self-selected news stories got slightly better ratings than the stories the researchers had chosen for the test. This indicates that personalized news stories have higher end-user value (similarly, see Kim & Lee, 2019). In addition, the result can tentatively be interpreted as meaning that people will tolerate some linguistic imperfections if the topic is interesting enough and they are part of the dialogue.

While this is good news for those involved in the pursuit of human-like linguistic fluency, it leads us onto an endless path of compromise between idealism and reality. It also forces us to consider a fundamental question: does it make sense in the first place that an NLG-driven system featuring a high level of automation should aim to mimic journalism written by humans? It looks like a race that NLG is destined to lose, because the genre of *human-written* journalism has adapted and evolved in accordance with *human* affordances. A reasonable analogy would be the juxtaposition of a breaking news alert and a piece of investigative journalism: it makes little sense to compare their fluency and depth of consideration as they are not even attempting to fill the same need, but rather two complementing needs of the audience.

On this basis, we argue that a more feasible path to take to improve NLG-driven news automation would be to create a genre on its own terms by focusing consciously on the strengths of NLG – as we did with personalization. It should be noted that these advantages go beyond the simple speed of the automated system.

4. *Valtteri* does not store the reader data entered in its user interface. Deeper modelling based on the online and offline behaviour of the reader would have to take into account the strict regulation introduced by the General Data Protection Regulation (GDPR).

For example, credibility and aspects close to it such as equality and objectivity were perceived as advantages in the user test. Perhaps it makes sense to focus development on domains and topics that are prone to value-laden biases and doubts about them, an obvious example being politics and elections. As also shown by the user test, the impression of credibility can be achieved through simple, straightforward language rather than any great linguistic virtuosity, and in addition to running text, the use of numbers, tables, mind maps and lists would clarify and illustrate the issue in question while avoiding linguistic pitfalls (as for *genre blending*, see Mäntynen & Shore, 2014). As our research and development team concluded elsewhere (Lindén & Tuulonen et al., 2019, p. 30):

You can often transform dry-sounding numbers into invigorating stories if you can only find the relevant connection to the reader. That connection may be places or people of interest, statistical trends, and abnormalities, or relation to popular trivia – to name but a few.

Furthermore, the high computing power that enables this fantastic speed and personalization could also be harnessed for journalistic use in other ways. For example, algorithms are better equipped than people to find hidden relationships, weak signals, and outliers, and the triangulation of several data sets would further help to make the most out of the computing power. Naturally, this also brings with it many challenges – especially in terms of credibility – that further research needs to address.

5. Conclusion

In this paper, we have explained how we recontextualized the determinant aspects of a genre of online news story in the algorithm of our NLG-driven news automation software. We then went through the key results of the user test, which were generally unflattering but indicate some potential for further improvement of the software, and concluded by questioning the idea that NLG-driven systems should mimic journalism written by humans. Instead, we suggested that development work in the field of news automation should aim to create a new genre that is based on the inherent strengths of NLG, and we presented some tentative ideas as to what this genre could include. However, it remains for further research and practical applications to figure out what the characteristics of this genre could be, and how they could become established.

At the very beginning we were motivated by the fact that the amount of available data is increasing at a tremendous rate, but the data needs to be converted into a user-friendly form in order to be made use of. However, while the techno-

logical development of NLG systems is happening fast and is producing exciting results in many fields of society, and while the media business too has a variety of algorithmic applications in the gathering, production and distribution of news (Beckett, 2019; Diakopoulos, 2019), the progress of NLG-driven news automation has been steady but slow (Hansen et al., 2017; Lindén, 2017; Lindén & Tuulonen et al., 2019). Regardless of the technical difficulties of applying NLG in journalism, some of which we have raised in this paper, it is very much the data that continue to be the most serious bottleneck: so far, the domains available for automation are limited to such fields as sports and election results, key financial figures, and traffic and weather conditions, where the data consist of numbers or other known values and the range of possible news stories is well understood. Sirén-Heikel et al. (2019), in interviews with representatives of the media, came across similar views on the lack of both technological and data quality, as well as another issue, a lack of expertise in newsrooms.

When all is said and done, it is important not only that we should continue to work on the technical development of NLG, but also that evaluations of the usefulness of the technology should not be burdened by our preconceptions of what news is and what a news story should look like. Instead, it should be recognized that automated journalism can be viewed as its own, unique, subgenre, with its own distinct strengths, weaknesses, and requirements. Only after that, we argue, can NLG-driven systems featuring a high level of automation really benefit journalistic practice outside of niche applications.

We hypothesize that a careful analysis of the types of domains where news automation has been successfully applied so far would reveal that the prototypical texts in these domains are closer in nature to the new type of genre we have sketched above than to the prototypical exemplar of investigative journalism. This is where the future lies.

Funding

This paper is supported by the European Union's Horizon 2020 research and innovation programme under grant agreement No 825153, project EMBEDDIA (Cross-Lingual Embeddings for Less-Represented Languages in European News Media).

References

- Beckett, C. (2019). New powers, new responsibilities. A global survey of journalism and artificial intelligence. *The Journalism AI*. Retrieved from <<https://blogs.lse.ac.uk/polis/2019/11/18/new-powers-new-responsibilities>> (22 February, 2020).
- Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. *ACM Computing Surveys*, 41(3), 15:1–15:58. <https://doi.org/10.1145/1541880.1541882>
- Clerwall, C. (2014). Enter the robot journalist: Users' perceptions of automated content. *Journalism Practice*, 8(5), 519–531. <https://doi.org/10.1080/17512786.2014.883116>
- Devitt, A. J. (2004). *Writing genres*. Carbondale, IL: Southern Illinois University Press.
- Diakopoulos, N. (2019). *Automating the news. How algorithms are rewriting the media*. Cambridge, MA: Harvard University Press. <https://doi.org/10.4159/9780674239302>
- Fairclough, N. (1992). *Discourse and social change*. Cambridge: Polity Press.
- Gatt, A., & Krahmer, E. (2017). Survey of the state of the art in natural language generation: Core tasks, applications and evaluation. *Journal of Artificial Intelligence Research*, 60, 75–170.
- Graefe, A., Haim, M., Haarmann, B., & Brosius, H.-B. (2016). Readers' perception of computer-generated news: Credibility, expertise, and readability. *Journalism*, 19(5), 595–610. <https://doi.org/10.1177/1464884916641269>
- Gruber, H. (2019). Genres, media, and recontextualization practices: Re-considering basic concepts of genre theory in the age of social media. *Internet Pragmatics*, 2(1): 54–82. <https://doi.org/10.1075/ip.00023.gru>
- Gupta, M., Gao, J., Aggarwal, C. C., & Han, J. (2014). Outlier detection for temporal data: A survey. *IEEE Transactions on Knowledge and Data Engineering*, 26(9), 2250–2267. <https://doi.org/10.1109/TKDE.2013.184>
- Hansen, M., Roca-Sales, M., Keegan, J. M., & King, G. (2017). Artificial intelligence: Practice and implications for journalism. *Columbia University Academic Commons*. <https://doi.org/10.7916/D8X92PRD>
- Kim, D. & Lee, J. (2019). Designing an algorithm-driven text generation system for personalized and interactive news reading. *International Journal of Human-Computer Interaction*, 35(2), 109–122. <https://doi.org/10.1080/10447318.2018.1437864>
- Latar, N. L. (2015). The robot journalist in the age of social physics: The end of human journalism? In G. Einav (Ed.), *The new world of transitioned media: Digital realignment and industry transformation* (pp. 65–80). Wiesbaden: Springer. https://doi.org/10.1007/978-3-319-09009-2_6
- Lee, A. M. (2014). How fast is too fast? Examining the impact of speed-driven journalism on news production and audience reception (Unpublished doctoral dissertation). The University of Texas at Austin.
- Leppänen, L., Munezero, M., Granroth-Wilding, M., & Toivonen, H. (2017a). Data-driven news generation for automated journalism. In *Proceedings of the 10th International Conference on Natural Language Generation*, 188–197. <https://doi.org/10.18653/v1/W17-3528>
- Leppänen, L., Munezero, M., Sirén-Heikel, S., Granroth-Wilding, M., & Toivonen, H. (2017b). Finding and expressing news from structured data. In *Proceedings of the 21st International Academic Mindtrek Conference*, 174–183. <https://doi.org/10.1145/3131085.3131112>

- Lindén, C.-G. (2017). Decades of automation in the newsroom: Why are there still so many jobs in journalism? *Digital Journalism*, 5(2), 123–140.
<https://doi.org/10.1080/21670811.2016.1160791>
- Lindén, C.-G., & Tuulonen, H. (Eds.) together with Bäck, A., Diakopoulos, N., Haapanen, L., Leppänen, L., Melin, M., Munezero, M., Sirén-Heikel, S., Södergård, C., & Toivonen, H. (2019). *News Automation: The rewards, risks and realities of “machine journalism”*. WAN-IFRA guide to the field. Reports / The World Association of Newspapers and News Publishers WAN-IFRA.
- Linell, P. (1998). *Approaching dialogue. Talk, interaction and contexts in dialogical perspectives*. Amsterdam: John Benjamins. <https://doi.org/10.1075/impact.3>
- Luginbühl, M. (2014). Genre profiles and genre change: The case of TV news. In J. Androustopoulos (Ed.), *Mediatization and Sociolinguistic Change* (pp. 305–330). Berlin, New York: de Gruyter. <https://doi.org/10.1515/9783110346831.305>
- Martin, J.R. (1985). Process and text: two aspects of human semiosis. In J.D. Benson, & W.S. Greaves (Eds.), *Systemic perspectives on discourse* (pp. 248–274). Norwood, NJ: Ablex.
- Melin, M., Bäck, A., Södergård, C., Munezero, M., Leppänen, L., & Toivonen, H. (2018). No landslide for the human journalist. An empirical study of computer-generated election news in Finland. *IEEE Access*, 6, 43356–43367. <https://doi.org/10.1109/ACCESS.2018.2861987>
- Miller, C. (1984). Genre as social action. *Quarterly Journal of Speech* 70(2), 151–167.
<https://doi.org/10.1080/00335638409383686>
- Miller, C., & Shepherd, D. (2009). Questions for genre theory from the blogosphere. In J. Giltrow (ed.), *Genres in the Internet: Issues in the theory of genre* (pp. 264–286). Amsterdam: John Benjamins. <https://doi.org/10.1075/pbns.188.11mil>
- Murphy, K.P. (2012). *Machine learning: A probabilistic perspective*. Cambridge, MA: The MIT press.
- Mäntynen, A., & Shore, S. (2014). What is meant by hybridity? An investigation of hybridity and related terms in genre studies. *Text and talk*, 34(6), 737–758.
<https://doi.org/10.1515/text-2014-0022>
- O'Neill, D., & Harcup, T. (2009). News values and selectivity. In Wahl-Jorgensen, K. & Hanitzsch, T. (Eds.) *Handbook of journalism studies* (pp. 161–174). New York, NY: Routledge.
- Pietikäinen, S., & Mäntynen, A. (2020). *Uusi kurssi kohti diskurssia*. Tampere: Vastapaino.
- Pöttker, H. (2003). News and its communicative quality: the inverted pyramid – when and why did it appear? *Journalism Studies*, 4(4), 501–511.
<https://doi.org/10.1080/1461670032000136596>
- Rosenberg, H., & Feldman, C.S. (2008). *No time to think: The menace of media speed and the 24-hour news cycle*. New York, NY: Continuum.
- Weaver, D.H., & Willnat, L. (Eds.). (2012). *The global journalist in the 21st century*. London: Routledge. <https://doi.org/10.4324/9780203148679>
- Wölker, A., & Powell, T.E. (2018). Algorithms in the newsroom? News readers' perceived credibility and selection of automated journalism. *Journalism*.
<https://doi.org/10.1177/1464884918757072>
- Zampa, M. (2017). *Argumentation in the newsroom*. Amsterdam: John Benjamins.
<https://doi.org/10.1075/aic.13>

Sirén-Heikel, S., Leppänen, L., Lindén, C.-G., & Bäck, A. (2019). Unboxing news automation: Exploring imagined affordances of automation in news journalism. *Nordic Journal of Media Studies* 1(1), 47–66. <https://doi.org/10.2478/njms-2019-0004>

Address for correspondence

Lauri Haapanen
University of Jyväskylä
P.O. Box 35, FI-40014
Finland
lauri.m.haapanen@jyu.fi

Co-author information

Leo Leppänen
University of Helsinki
Department of Computer Science
leo.leppanen@helsinki.fi